Poster I-49

RNA Secondary Structure Prediction With Simple Pseudoknots

Deogun, Jitender S., Donis, Ruben, Komina, Olga, Ma, Fangrui

Pseudoknots are widely occurring structural motifs in RNA. Pseudoknots have been shown to be functionally important in different RNAs which play regulatory, catalytic, or structural roles in cells. Current biophysical methods to identify the presence of pseudoknots are extremely time consuming and expensive. Therefore, bioinformatic approaches to accurately predict such structures are highly desirable.

Most methods for RNA folding with pseudoknots approaches follow like quasi-Monte Carlo search, genetic algorithms, stochastic context-free grammars, and the Hopfield networks, and dynamic programming (DP). These approaches, however, have limitations. The DP algorithm has worst case time and space complexities of $O(n^{6.8})$ and $O(n^4)$, respectively. The algorithm is not practical for sequences longer than 100 nucleotides.

In this paper, we present a dynamic programming algorithm for predicting optimal secondary structure including simple pseudoknots for single RNA sequences using standard thermodynamic parameters for RNA folding. Our approach is based on a pseudoknot technique for maximizing the number of base pairs. The algorithm has worst case time and space complexities of $O(n^4)$ and $O(n^3)$, respectively.

We validate the accuracy of our algorithm by experimental results on the entire set of simple pseudoknot collection in the PseudoBase. Our program folds 160 pseudoknots out of 169 totals in the PseudoBase database predicting the structure of 131 pseudoknot correctly or almost correctly. The algorithm is quite efficient. For example, a sequence of 75 nucleotides takes 55 seconds (compared to 20 minutes with existing software) and a sequence of 114 nucleotides takes 8 minutes (4 hours 30 min). To our knowledge, this is most accurate and efficient algorithm for predicting optimal secondary structure of a single RNA sequence including simple pseudoknots.